THE SIR WILLIAM MACLEAY MEMORIAL LECTURE 1989

Haemoglobin in Plants: Evolution really is Conservative

JIM PEACOCK
Division of Plant Industry, CSIRO, Canberra
(Delivered by J. PEACOCK 11 October 1989)

It is a pleasure for me to be giving the Linnean MacLeay Lecture, particularly because I was a Linnean MacLeay Fellow during my PhD program in this University. At that time my research concerned the evolution of the Australian flora. I was concentrating on one family, the *Goodeniaceae*, under the guidance of Professor Spinney Smith-White whom I am sorry is not able to be here tonight.

One of the critical tools we used in making deductions about relationships between species and genera within the plant family was the number of chromosomes in the genome. I can remember looking at countless chromosome spreads and wishing that I could be down inside the nucleus looking at gene activity, thinking that if I could really look at genes I'd be able to say much more about the relationships of species and how evolution really worked.

That may have been a naive dream at the time but what I'd like to discuss this evening is the achievement of that dream. These days we are able to look at genes; not only at their genetic structure and the map of their genetic code but also we are having some considerable success in understanding how their activity is controlled.

Controlled patterns of gene activity are really the major determinants of development.

I am privileged to work with some wonderful colleagues in CSIRO, and we and many other laboratories working in this gene engineering field have been struck by the power of selection and the conservation of successful molecular strategies over huge distances within the evolutionary fabric.

The examples I am about to give you on gene sequence and structure and of the systems which control gene activity are comparable to the morphological and physiological characters that have taught us so much about evolution in the past. But to me the new observations are more emphatic because we are looking at properties of the prime genetic material. My examples are from our own laboratory but I want to emphasise that much of what I have to say is paralleled and often depends upon work in many other laboratories around the world.

One example I think you will find interesting concerns the molecule which gives our blood its characteristic red colour, haemoglobin. Haemoglobin is important in many animals for oxygen transport and it is also found in plants. Until fairly recently it was known only in nodules on the roots of legumes. These are structures that are formed in a symbiotic relationship between the bacterium *Rhizobium* and the host plant root tissue. They are important for the fixation of atmospheric nitrogen for the plant. I hasten to add, because this is a Sydney audience, that the plant that you know in our own flora, *Hemadorum*, with its brilliant red tissue, is not a plant with a lot of haemoglobin; the orange red colour is another pigment molecule.

The haemoglobin in nodules, where does it come from? It is produced by the plant tissue. Initially, when the amino acid sequence was determined, it appeared that

there were striking similarities between the legume haemoglobin molecule and the haemoglobin in animals. In particular, critical amino acids involved in folding of the molecule, in forming bonds with the haem moiety, and in forming the O_2 pocket, were identical in the plant and animal molecules. But it was when the plant haemoglobin gene was analysed that really striking identity emerged.

As you may know, genes in plants and animals often have an interrupted genetic code structure, with alternating exons and introns. There are three introns in the leghaemoglobin gene. The first and third introns are in precisely the same positions as the two introns in animal haemoglobins. The central intron is at the spot where protein chemists had predicted from the domain structure of the animal haemoglobin that another intron could have been expected. This was on the basis of the supposition that introns are involved with the construction of genes from component sequences. This lego theory of gene construction has considerable support.

It was realised that this identity was not likely to have been by chance. There were basically two explanations. Both assume that the genes were related by evolutionary descent. In one case we can assume a linear descent from a progenitor of plants and animals. On the other hand we can propose transfer of the gene from the animal kingdom into the legumes during the evolution of the symbiotic relation between the bacteria and the legume plant. Insects were considered to be a possible donor.

The plot thickened when the haemoglobin molecule was found in other plants, plants which have nodules that fix nitrogen — *Casuarina* is a good example, and another one is *Parasponia* which is a member of the elm family. There are many other examples known, scattered throughout the dicotyledons. Initially haemoglobin was not detected, but subsequently with better biochemistry, it has been shown that haemoglobin is present in the molecules in each species.

A couple of years ago we isolated the haemoglobin genes, initially from *Parasponia* and then later from *Casuarina*. Their sequence and intron positions left little doubt that all plant haemoglobin genes are closely related. This obviously placed some strain on the horizontal evolution hypothesis of haemoglobins into plants and strongly favoured the linear descent concept. In thinking about this we wondered whether this meant, since the phylogenetic distribution of nodulated plants is so scattered, that previously all plants were nitrogen-fixing and were nodulated, with only a few species in several families being the relics of this condition. Or could it mean that in fact all plants have a haemoglobin gene, presumably with a function or functions in a normal plant. Particular plants would have evolved a symbiotic relation with *Rhizobium* or with *Franckia* or some other bacteria independently in different families or groups of families. This latter alternative is much more attractive in an evolutionary sense and on a lot of morphological grounds too with respect to the structure of nodules and the biology of the symbiotic relationships.

We then looked at some plants that were known not to nodulate and asked whether they have a haemoglobin gene present and in a functioning form. It is difficult to look for the molecule if it is in very low amounts. The isolation of haemoglobin is no simple task from plant tissue. There are antibodies for haemoglobin but the trouble is that haemoglobin, although having certain key regions highly conserved, is a protein which permits a substantial amount of amino acid substitution, presumably without disrupting functional aspects of the molecule. The antibody is likely then to have limited phylogenetic reach. The antibody will only react where proteins are fairly closely related and have the same epitopes.

Similarly, we would expect that nucleic acid probes would have a limited taxonomic reach too. The amino acid substitution variation means that we can expect a lot of mutation substitution in the DNA sequences, and this is the case. Initially we started

with a particular family. We chose the elm family where both nodulated and non-nodulated species and genera were known. We found that *Trema*, a genus closely related to *Parasponia* but which does not nodulate, does have a haemoglobin gene and it is a functioning gene. So too does another genus, *Celtis*, and we think that *Ulmus* itself also has the gene. On this basis we feel it is likely that all plants have a haemoglobin gene.

We asked first in *Trema* where the gene did function in the plant. Initially we checked stem, root and leaves and found expression, detecting both messenger RNA and protein, only in the roots. The fact that there was tissue-specific expression suggested that this haemoglobin must have a particular function or functions, hitherto unsuspected in plants. We have not looked extensively in other tissues but there has been a report of haemoglobin in the seed of the winged bean. In *Parasponia*, the single haemoglobin gene is expressed in large amount in the nodules and to a much lower level in the roots. There must be two different controls operating.

We took our analysis a little further by making transgenic plants, introducing the *Trema* and the *Parasponia* genes into tobacco. We found that the gene was expressed in the roots of the transgenic tobacco plant. The genes, when introduced into *Lotus*, a legume, were expressed in both the nodule and in the root.

At this stage I need to say a little bit more about the structure of a gene. As well as the coding region which determines the gene product, another important property that has become clear in recent years is that there is an upstream region in the DNA molecule, immediately adjacent to the coding region, which contains most of the signals important in determining the time, place and amount of gene expression.

When we placed the *Parasponia* and *Trema* genes into tobacco, or even just their control regions hooked to a reporter gene, which enabled us to determine easily where the gene was working, the results indicated that the gene controls, which evolved in the elm family of plants, were working perfectly well in the transgenic legume and in transgenic tobacco.

There are two components in an effective gene control element; the DNA target sequence and a binding protein. Our results implied that the tobacco plant has the machinery to correctly control the expression of the *Ulmaceae* haemoglobin gene; presumably it has its own haemoglobin gene, highly homologous to those we inserted by genetic engineering. Our conclusion is that almost certainly the primitive condition in plants is that there is a haemoglobin gene in the genome and that the gene product has one or more functions in certain plant tissues. We suspect that this is probably the case in all plants.

What could that function be? Initially, we judged by the very small amount of haemoglobin present in roots, that unless there was localization to particular cells there was not enough haemoglobin present for it to act as an oxygen carrier, the same way it does in nodules. We wondered whether it might be an oxygen sensor molecule. Using the reporter genes that I talked about before we found that expression is localized to particular cells in the root, so the possibility that it acts as an oxygen carrier is still a viable alternative. The take-home message is that the conservatism of evolution is beautifully shown by the haemoglobin genes; it is highly probable that both the animal and plant kingdom variants have evolved from a common ancestor organism which preceded the separate evolutionary paths of the animal and plant kingdoms. Recent data suggest that the gene may extend back to proto-organisms occurring some 3-5 billion years ago.

I want to push these molecular examples a little further now by giving you another story from our lab, concerning gene controls. I think the conservatism of controls of gene action provides perhaps the most striking evidence of the powers of selection and the opportunism of evolution that I have seen in molecular analyses of the genome. We

have been studying an important biochemical response in plants. When roots are flooded, oxygen is excluded from the cell environment and plants switch their metabolism from an oxidative breakdown of carbohydrate to a fermentative pathway; there is a remarkable cessation of protein synthesis followed by a selective synthesis of anaerobic proteins, which have been identified as the enzymes of the fermentation pathway. One of the critical enzymes is alcohol dehydrogenase (Adh), now probably the best known gene system in plants. We chose originally to study it, not only because of its presence in this particular facet of plant biochemistry, but because there were some strong genetic tools available to us. The alcohol dehydrogenase gene structure, quite a complex one with nine introns, is conserved through the plant kingdom.

With regard to the control of this enzyme, what we did was to do some genetic surgery in the upstream region, making deletions, and then putting the deleted gene back into plant cells and asking whether it could still work in the right way. We were able to define a small, critical region that donated the property of anaerobic response in the transcription of this particular gene. If we took that small region, the anaerobic response element — the ARE — and put it in front of another gene then the target gene came under anaerobic induction control. The ARE has a highly conserved core hexamer, TGGT T T, and we found this present in the critical ARE regions of every other alcohol dehydrogenase gene we have looked at in both monocot and dicot species. Interestingly it is also present in the control regions of the other genes that are under anaerobic control, for example in the aldolase and sucrose synthase genes. This is an example of where a particular control region has been used to bring about coordinate expression of genes in order to achieve a complex biochemical response.

The aldolase gene also demonstrates a point I mentioned earlier, that genes have been put together in the way that a child builds a lego construction. We suspect that maize aldolase provides us with an example of an upstream control region being hooked to an existing aldolase sequence in order to give it an anaerobic induction property.

This same anaerobic control we characterized in plants is now being found in anaerobically controlled genes in fungi and even in bacteria. A wonderful example of the way in which selection has maintained something that has the right properties. The striking thing to me is that selection can be so tight that it preserves a sequence of just six nucleotides.

But, I want to stress that the conservatism isn't just in DNA sequences. Remember I mentioned before that the control sequence that we are looking at is a target sequence. DNA binding proteins recognize particular target sequences and it is the combination of the bound protein and the DNA sequence that provides the control switch. In fact there are usually a number of control components which interact to give an effective transcription unit to provide the signal for the RNA polymerase molecule to proceed down the DNA double helix transcribing a messenger RNA. In probing the anaerobic induction control of Adh, we asked whether the maize control would work in transgenic tobacco. We were initially disappointed to find that it hardly worked at all. But when we added an enhancer sequence, a sequence which increased the amplitude of gene response, we found the maize ARE worked fine in tobacco. In this case the combination of the enhancer sequence and the ARE were necessary to give a high intensity and well controlled response.

I have mentioned this because we have recently put a lot of effort into examining the protein that binds to the enhancer sequence. Of course this DNA binding protein is itself encoded by a gene. Our surprise was that when we isolated and sequenced the gene, we found, in the deduced protein sequence, extreme similarity to the amino acid sequences of binding proteins known in animals and in yeast. There is a particular class of DNA binding protein called a leucine zipper. The name comes from a regular spacing

of leucine residues in part of the protein molecule concerned with dimerization — two polypeptides associate together to form a dimer which is the functional protein. It is thought that the leucines interact in much the same way as two sides of a zipper. Adjacent to the leucine zipper region in the protein is a region of basic residues which is the region that recognizes and binds to the target DNA sequence. Elsewhere in the protein, in some proteins towards the N terminus and in others towards the carboxy terminus, there is a region, the activating region, involved in the interactions to make the effective transcription unit.

The remarkable thing is that many genes in the animal, plant and fungal kingdoms have used and maintained, with an extraordinarily high degree of conservation, this control system. It is a beautiful example of co-evolution, with conservation of both the DNA target sequence and the sequence of amino acids in the protein, important in bringing about the DNA-protein interaction. The system has been used in the different organisms to control different classes of genes. The GCN4 case in yeast is common to many genes involved in amino acid biosynthesis pathways; c-JUN and other related proteins in animals are concerned with control of a number of basic cellular processes. In fact c-JUN is an oncogene — when it is under imperfect control itself it can lead to induction of cancers.

We as yet don't know how this particular enhancer sequence is used in plant metabolism. One thing we do know is that pathogens of plants, including both *Agrobacterium* and certain plant viruses, have trapped this highly efficient transcription control to use for their own purposes.

I have tried to give you a present-day look at one property of life. Even at the finest levels of examination of the primary genetic material, and of the control functions for that primary code, evolution has been innovative but conservative. These striking cases of evolutionary conservatism in gene coding sequences, control target sequences and control binding protein gene sequences are all compatible with a single origin of life forms. The story which is unfolding in an exciting and rapid way, is that fundamental controls of gene action underlie the intricate patterns of development in plants, animals and micro-organisms. Differential cell futures depend on differential control of transcriptional activity of genes. This is not the only mechanism involved in the control of development but it is one of the major ones. For further background on the topics discussed above I would refer you to Appleby *et al.* (1988a,b, 1989), Landsmann *et al.* (1988) and Peacock (1989).

The work plan of evolution is like the Button plan for Australia's car industry. Different lifeforms use and rely upon the same basic components!

References

Appleby, C. A., Bogusz, D., Dennis, E. S., and Peacock, W. J., 1988. — A role for haemoglobin in all plant roots? *Plant Cell and Environ.* 11: 359-67.

Appleby, C. A., Bogusz, D., Dennis, E. S., Fleming, A. I., Landsmann, J., and Peacock, W. J., 1988. — The vertical evolution of plant hemoglobin genes. *In* Bothe, H., De Bruijn, F. J., and Newton, W. E. (eds.) *Nitrogen Fixation: Hundred Years After.* New York. Gustav Fischer, Stuttgart.

Appleby, C. A., Dennis, E. S., and Peacock, W. J., 1989. — A primaeval origin for plant and animal haemoglobins. *Aust. J. Sys. Bot.* (in press).

Landsmann, J., Llewellyn, D., Dennis, E. S., and Peacock, W. J., 1988. — Organ regulated expression of the *Parasponia andersonii* haemoglobin gene in transgenic tobacco plants. *Mol. Gen. Genet.* 214: 68-73.

Peacock, W. J., 1989. — Johannsen's gene and genotype in the light of plant molecular biology. *Genome* 31: 33-41.